Faculty of Mathematical Sciences

University of Twente

University for Technical and Social Sciences

P.O. Box 217
7500 AE Enschede
The Netherlands
Phone: +31-53-4893400
Fax: +31-53-4893114
Email: memo@math.utwente.nl

Memorandum No. 1470

Generalized semi-infinite programming: Numerical aspects

G.J. Still

November 1998

ISSN 0169-2690

Generalized Semi-Infinite Programming: Numerical aspects

G. Still, University of Twente

Abstract

Generalized semi-infinite optimization problems (GSIP) are considered. It is investigated how the numerical methods for standard semi-infinite programming (SIP) can be extended to GSIP. Newton methods can be extended immediately. For discretization methods the situation is more complicated. These difficulties are discussed and convergence results for a discretization- and an exchange method are derived under fairly general assumptions on GSIP. The question is answered under which conditions GSIP represents a convex problem.

Keywords: Semi-infinite programming, numerical methods, discretization methods

Mathematical Subject Classification 1991: 90C34, 65K05, 90C30, 90C31

1 Introduction

We are concerned with *generalized semi-infinite optimization problems* of the form:

$$\text{GSIP:} \quad \min \ f(x) \qquad \text{subject to } \ x \in M = \{x \in I\!R^n \mid g(x,y) \geq 0, \ y \in Y(x)\}$$
$$\text{with} \quad Y(x) = \{y \in I\!R^r \mid v_l(x,y) \geq 0, \ l \in L\}$$

and L, a finite index set. If not stated otherwise, we assume that the functions f, g, v_l are C^2-functions and that the set valued mapping Y satisfies

$$Y \ : \ I\!R^n \to 2^{I\!R^r}, \quad Y(x) \subset C_0 \text{ for all } x \in I\!R^n \ \text{ with } \quad C_0 \subset I\!R^r \text{ compact.} \tag{1}$$

For the special case that the set $Y = Y(x)$ does not depend on x, i.e. $v_l(x,y) = v_l(y), \ l \in L$, the problem GSIP is a common semi-infinite problem and will be abbreviated by SIP. If moreover Y is a finite set then GSIP reduces to a finite optimization problem.

For a function $f(x)$ the derivative will be denoted by $Df(x)$ (row vector) and for a function $h(x,y)$ by $D_x h, \ D_y h$ we denote the partial derivatives w.r.t. the variables $x, \ y$. For brevity, we omit equality constraints in M and $Y(x)$.

Generalized semi-infinite problems have recently become a topic of interest. Optimality conditions for GSIP have been developed in [5], [6], [10], [12]. The structure of the feasible set has been investigated in [7], [11]. Some numerical aspects of GSIP are discussed in [12]. Numerical algorithms for a special class of GSIP (terminal problems, $y \in I\!R, r = 1$) are considered in [8]. In [9], GSIP's are studied with (in essence) functions $g(x,y) = \frac{1}{2}y^T G y + a^T y + y^T H x, \ G, H$, matrices, $v_l(x,y) = p_l^T y + q_l(x), p_l \in I\!R^r$ and convex functions $q_l, \ f$. By duality theory such a problem is reduced to a non-convex finite optimization problem. However, a general study of numerical methods for GSIP has not yet been done. With this paper we intend to make a first step.

For applications of GSIP in robotics (*maneuverability problem*), optimal control (*terminal problem*) and approximation theory (*reverse Chebyshev problem*) we refer to [3], [8] and [12].

The paper is organized as follows. In Section 2 the notation is introduced and optimality conditions based on 'local reduction' are given for later purposes. In Section 3 it is shown that the Newton-type methods can directly be generalized from SIP to GSIP. Section 4 is concerned with discretization- and exchange methods. The difference between the situation in SIP and GSIP is discussed. Convergence results for two types of algorithms are given under fairly natural assumptions. A discussion how these assumptions can be fulfilled in practice is done. A forthcoming paper will be concerned with numerical experiments on these algorithms. We do not consider so-called 'descent methods'. Section 5 investigates convex GSIP. Sufficient conditions are given for GSIP to represent a convex problem.

2 Preliminaries

In this section we give some preliminaries and outline optimality conditions for GSIP based on 'local reduction'. For $\overline{x} \in M$ we define the *set of active points*

$$Y_0(\overline{x}) = \{\overline{y} \in Y(\overline{x}) \mid g(\overline{x}, \overline{y}) = 0\} \ .$$

Obviously, for feasible $\overline{x} \in M$, any point $\overline{y} \in Y_0(\overline{x})$ is a (global) minimum of the following parametric optimization problem (the so-called *lower level problem*):

$$Q(\overline{x}): \qquad \min_{y} \quad g(\overline{x}, y) \quad \text{s.t.} \quad y \in Y(\overline{x}) \ . \tag{2}$$

Let in the sequel $v(x)$ denote the value function of $Q(x)$. Given $\overline{x} \in M$, $\overline{y} \in Y(\overline{x})$ we define the active index set $L_0(\overline{x}, \overline{y})$ with respect to $Q(\overline{x})$, $L_0(\overline{x}, \overline{y}) = \{l \in L \mid v_l(\overline{x}, \overline{y}) = 0\}$.

We say that at $\overline{y} \in Y(\overline{x})$ the '*linear independency constraint qualification*' (LICQ) is satisfied for $Q(\overline{x})$ if the vectors

$$D_y v_l(\overline{x}, \overline{y}), \quad l \in L_0(\overline{x}, \overline{y}) \ \text{ are linearly independent.} \tag{3}$$

The weaker '*Mangasarian Fromovitz constraint qualification*' (MFCQ) is said to hold at $\overline{y} \in Y(\overline{x})$ if

$$\text{there exists a vector } \eta \text{ such that } \ D_y v_l(\overline{x}, \overline{y})\eta > 0, \quad l \in L_0(\overline{x}, \overline{y}) \ . \tag{4}$$

Let be given $\overline{x} \in M$, $\overline{y} \in Y_0(\overline{x})$, i.e. $\overline{y}$ is a solution of $Q(\overline{x})$. If at $\overline{y}$ the MFCQ is satisfied then, necessarily the following Kuhn-Tucker condition is fulfilled: There exists a multiplier vector $\overline{\gamma} \in I\!R^{|L_0(\overline{x}, \overline{y})|}$ such that

$$D_y \mathcal{L}^{\overline{y}}(\overline{x}, \overline{y}, \overline{\gamma}) = 0 \ , \quad \overline{\gamma} \geq 0 \quad \text{with} \quad \mathcal{L}^{\overline{y}}(x, y, \gamma) = g(x, y) - \sum_{l \in L_0(\overline{x}, \overline{y})} \gamma_l v_l(x, y), \tag{5}$$

the Lagrange function. The following F. John type optimality condition holds for GSIP (cf. [10] for a short proof).

Theorem 1 *Let be given $\overline{x} \in M$. Suppose, at any point $\overline{y} \in Y_0(\overline{x})$ the MFCQ is satisfied for $Q(\overline{x})$. Then, there exist $\overline{y}^j \in Y_0(\overline{x})$, $\overline{\gamma}^j \in I\!R^{|L_0(\overline{x}, \overline{y}^j)|}$, $\overline{\gamma}^j \geq 0$, $j = 1, \ldots, p$, and multipliers $\overline{\mu}_0, \overline{\mu}_1, \ldots \overline{\mu}_p \geq 0$, not all zero, such that*

$$\overline{\mu}_0 Df(\overline{x}) - \sum_{j=1}^{p} \overline{\mu}_j D_x \mathcal{L}^{\overline{y}^j}(\overline{x}, \overline{y}^j, \overline{\gamma}^j) = 0 \ . \tag{6}$$

If $Y_0(\overline{x}) = \{\overline{y}^1, \ldots, \overline{y}^p\}$ and LICQ is satisfied at $\overline{x}$ (for GSIP), i.e.

$$D_x \mathcal{L}^{\overline{y}^j}(\overline{x}, \overline{y}^j, \overline{\gamma}^j), \ j = 1, \ldots, p, \quad \textit{are linearly independent} \tag{7}$$

then, we can assume $\overline{\mu}_0 = 1$ (Kuhn-Tucker condition) and the multipliers $\overline{\mu}_1, \ldots, \overline{\mu}_p$ are uniquely determined. Note, that for SIP the functions $v_l = v_l(y)$ do not depend on x. Consequently, $D_x \mathcal{L}^{\overline{y}^j}(\overline{x}, \overline{y}^j, \overline{\gamma}^j) = D_x g(\overline{x}, \overline{y}^j)$ in this case and (6) takes the form

$$\overline{\mu}_0 Df(\overline{x}) - \sum_{j=1}^{p} \overline{\mu}_j D_x g(\overline{x}, \overline{y}^j) = 0 \, . \tag{8}$$

For later purposes, we summarize second order optimality conditions for GSIP (cf. [5], [12] for proofs and details). Standard assumptions for the so-called 'reduction ansatz' to obtain second order conditions are the following: Let at any active point $\overline{y}^j \in Y_0(\overline{x})$ condition (3) hold and (5) with $\overline{\gamma}^j > 0$ (*strict complementary slackness*) as well as the second order conditions,

$$\eta^T D_y^2 \mathcal{L}^{\overline{y}^j}(\overline{x}, \overline{y}^j, \overline{\gamma}^j)\eta > 0, \quad \text{for all } \ \eta \in T(\overline{x}, \overline{y}^j) \setminus \{0\} \, , \tag{9}$$

where $T(\overline{x}, \overline{y}^j) = \{\eta \in I\!R^r \mid D_y v_l(\overline{x}, \overline{y}^j)\eta = 0, \quad l \in L_0(\overline{x}, \overline{y}^j)\}$. In the following we put $v^j := [v_l, \ l \in L_0(\overline{x}, \overline{y}^j)]^T$ (a matrix with rows v_l). The implicit function theorem applied to the system

$$D_y \mathcal{L}^{\overline{y}^j}(x, y^j, \gamma^j) = 0, \qquad v^j(x, y^j) = 0 \tag{10}$$

implies the existence of C^1-functions $y^j(x), \ \gamma^j(x)$ defined on a neighborhood $U(\overline{x})$ of $\overline{x}$ such that on $U(\overline{x})$ the value $y^j(x)$ is a local solution of $Q(x)$ with corresponding multiplier vector $\gamma^j(x)$ satisfying $y^j(\overline{x}) = \overline{y}^j, \ \gamma^j(\overline{x}) = \overline{\gamma}^j$. By implicitly differentiating (10) w.r.t. x we find the following formula for $Dy^j, \ D\gamma^j$,

$$\begin{array}{lcl} -D_{xy}\mathcal{L}^{\overline{y}^j}(\overline{x}, \overline{y}^j, \overline{\gamma}^j) & = & D_y^2 \mathcal{L}^{\overline{y}^j}(\overline{x}, \overline{y}^j, \overline{\gamma}^j) \ Dy^j(\overline{x}) - D_y^T v^j(\overline{x}, \overline{y}^j) \ D\gamma^j(\overline{x}) \\ -D_x v^j(\overline{x}, \overline{y}^j) & = & D_y v^j(\overline{x}, \overline{y}^j) \ Dy^j(\overline{x}) \end{array} \tag{11}$$

The assumptions (3) and (9) imply that the matrices (Jacobian of (10) w.r.t. y, γ)

$$\overline{M}^j := \begin{pmatrix} D_y^2 \mathcal{L}^{\overline{y}^j}(\overline{x}, \overline{y}^j, \overline{\gamma}^j) & -D_y^T v^j(\overline{x}, \overline{y}^j) \\ D_y v^j(\overline{x}, \overline{y}^j) & 0 \end{pmatrix} \qquad \text{are regular.} \tag{12}$$

Moreover, these conditions imply that the set $Y_0(\overline{x})$ is finite, $Y_0(\overline{x}) = \{\overline{y}^1, \ldots, \overline{y}^p\}$. Under these strong assumptions the problem GSIP can locally, in a neighborhood $U(\overline{x})$ of $\overline{x}$, be transformed into the following equivalent finite optimization problem (*reduced problem*):

$$\text{GSIP}_{\text{loc}}(\overline{x}) : \quad \min \ f(x) \quad \text{ s.t. } \ g^j(x) := g(x, y^j(x)) \geq 0, \quad j = 1, \ldots, p.$$

Here, the functions $y^j(x)$ are the local solutions of $Q(x)$ constructed above. By applying optimality conditions of finite optimization to the problem $\text{GSIP}_{\text{loc}}(\overline{x})$ we obtain the following sufficient optimality conditions for GSIP (cf. e.g. [12]): Let at all points in $Y_0(\overline{x}) = \{\overline{y}^1, \ldots, \overline{y}^p\}$ the above standard assumptions be satisfied. Assume that at $\overline{x} \in M$

the condition LICQ is fulfilled (cf. (7)), as well as the Kuhn-Tucker condition (i.e. (6) holds with $\overline{\mu}_0 = 1$) and the second order condition,

$$\xi^T\, \overline{M}_0\, \xi > 0 \quad \text{for all } \xi \in T \setminus \{0\} \tag{13}$$

where $T = \{\xi \in I\!R^n \mid D_x \mathcal{L}^{\overline{y}^j}(\overline{x}, \overline{y}^j, \overline{\gamma}^j)\xi = 0,\ j = 1, \ldots, p\}$ and

$$\begin{aligned} \overline{M}_0 \ &:= \ \overline{\mu}_0 D^2 f(\overline{x}) - \sum_{j=1}^{p} \overline{\mu}_j D_x^2 \mathcal{L}^{\overline{y}^j}(\overline{x}, \overline{y}^j, \overline{\gamma}^j) + \sum_{j=1}^{p} \overline{\mu}_j D^T y^j(\overline{x}) D_y^2 \mathcal{L}^{\overline{y}^j}(\overline{x}, \overline{y}^j, \overline{\gamma}^j) D y^j(\overline{x}) \\ &\quad + \sum_{j=1}^{p} \overline{\mu}_j \sum_{l \in L_0(\overline{x}, \overline{y}^j)} \overline{\gamma}_l^j \Big(D^T \gamma_l^j(\overline{x}) D_x v_l(\overline{x}, \overline{y}^j) + D_x^T v_l(\overline{x}, \overline{y}^j) D \gamma_l^j(\overline{x}) \Big) \end{aligned} \tag{14}$$

Then, $\overline{x}$ is a local minimizer of GSIP.

We end up this section with short comments on the difference between SIP and GSIP. Under the standard assumptions above, for SIP the feasible set $M = \{x \in I\!R^n \mid g(x, y) \geq 0,\ \forall y \in Y\}$ is always closed. For GSIP this need not be the case (see Example 2 and [6, Section 2], [12]). Another phenomenon in GSIP is, that even if MFCQ is satisfied at any point $\overline{y} \in Y(\overline{x})$, the feasible set M of GSIP may have 're-entrant corners' at $\overline{x}$. We refer to [10] and [12] for examples and further details. This behavior is excluded if LICQ is satisfied for $Q(\overline{x})$ at all points $\overline{y} \in Y(\overline{x})$ (cf. [12, Theorem 3]).

3 Newton's method for solving GSIP

A common method for solving SIP is to apply Newton's method (or a Quasi-Newton variant) to the necessary optimality conditions (see e.g. [1], [4]). In [12] it is indicated that this approach can directly be generalized from SIP to GSIP. Here, we will give a proof of this assertion under the 'standard assumptions' in Section 2.

Consider $\overline{x} \in M$ such that at any point $\overline{y}^j \in Y_0(\overline{x}),\ j = 1, \ldots, p$, the conditions (3), (9) are satisfied. Let moreover (7) and (13) be fulfilled. Then, necessarily (cf. Theorem 1) $\overline{x}, \overline{\mu}, \overline{y}^j, \overline{\gamma}^j,\ j = 1, \ldots, p$, will solve the following system of Karush-Kuhn-Tucker equations of GSIP and the corresponding lower level problem $Q(\overline{x})$:

$$\begin{aligned} Df(x) - \sum_{j=1}^{p} \mu_j \Big(D_x g(x, y^j) - \sum_{l \in L_0(\overline{x}, \overline{y}^j)} \gamma_l^j D_x v_l(x, y^j) \Big) \ &= \ 0 \\ g(x, y^j) - \sum_{l \in L_0(\overline{x}, \overline{y}^j)} \gamma_l^j v_l(x, y^j) \ &= \ 0 \qquad j = 1, \ldots, p \\ \text{and for } j = 1, \ldots, p \qquad & \\ D_y g(x, y^j) - \sum_{l \in L_0(\overline{x}, \overline{y}^j)} \gamma_l^j D_y v_l(x, y^j) \ &= \ 0 \\ v_l(x, y^j) \ &= \ 0 \qquad l \in L_0(\overline{x}, \overline{y}^j) \end{aligned} \tag{15}$$

This system consists of $K := n + p + \sum_{j=1}^{p}\big(r + |L_0(\overline{x}, \overline{y}^j)|\big)$ equations for the K unknowns $x \in I\!R^n,\ \mu_j \in I\!R,\ y^j \in I\!R^r,\ \gamma^j \in I\!R^{|L_0(\overline{x},\overline{y}^j)|}, j = 1, \dots, p$. In the following lemma it is proven that under our assumptions the Jacobian of the system (15) is regular at the solution. This in particular implies that the Newton method (Quasi-Newton method) applied to (15) will locally converge quadratically (super-linearly).

Lemma 1 *Let $\overline{x} \in M$ be given such that at any point $\overline{y}^j \in Y_0(\overline{x}),\ j = 1, \dots, p$ the conditions (3), (9) are satisfied and let (7), (13) be fulfilled. Then, the Jacobian of (15) at $\overline{x}, \overline{\mu}, \overline{y}^j, \overline{\gamma}^j,\ j = 1, \dots, p$, is regular.*

Proof. The Jacobian of the system (15) reads (all functions evaluated at $\overline{x}, \overline{\mu}, \overline{y}^j, \overline{\gamma}^j$):

$$\begin{pmatrix}
\overbrace{D^2 f - \sum_{j=1}^{p} \overline{\mu}_j D_x^2 \mathcal{L}^{\overline{y}^j}}^{x} & \overbrace{-B^T}^{\mu} & \overbrace{-\overline{\mu}_1 D_{yx}\mathcal{L}^{\overline{y}^1}}^{y^1} & \overbrace{\overline{\mu}_1 D_x^T v^1}^{\gamma^1} & \cdots & \overbrace{-\overline{\mu}_p D_{yx}\mathcal{L}^{\overline{y}^p}}^{y^p} & \overbrace{\overline{\mu}_p D_x^T v^p}^{\gamma^p} \\
B & 0 & 0 & 0 & \cdots & 0 & 0 \\
D_{xy}\mathcal{L}^{\overline{y}^1} & 0 & D_y^2 \mathcal{L}^{\overline{y}^1} & -D_y^T v^1 & \cdots & 0 & 0 \\
D_x v^1 & 0 & D_y v^1 & 0 & \cdots & 0 & 0 \\
\vdots & & & & \ddots & & \vdots \\
D_{xy}\mathcal{L}^{\overline{y}^p} & 0 & 0 & 0 & \cdots & D_y^2 \mathcal{L}^{\overline{y}^p} & -D_y^T v^p \\
D_x v^p & 0 & 0 & 0 & \cdots & D_y v^p & 0
\end{pmatrix} \tag{16}$$

where $B^T := [D_x^T \mathcal{L}^{\overline{y}^1}, \dots, D_x^T \mathcal{L}^{\overline{y}^p}]$ and in the rows $n+1, \dots, n+p$ we have used the relations $D_y \mathcal{L}^{\overline{y}^j} = 0,\ v^j = 0$. Now, for $j = 1, \dots, p$, we add to the first n columns of (16) a combination Dy^j of the columns corresponding to the variable y^j and a combination $D\gamma^j$ of the columns corresponding to the variable γ^j. Then, by using (11) and (12) the matrix (16) is transformed into the following matrix without changing the determinant,

$$\begin{pmatrix}
M_0 & -B^T & -\overline{\mu}_1 D_{yx}\mathcal{L}^{\overline{y}^1} & \overline{\mu}_1 D_x^T v^1 & \cdots & -\overline{\mu}_p D_{yx}\mathcal{L}^{\overline{y}^p} & \overline{\mu}_p D_x^T v^p \\
B & 0 & 0 & & \cdots & 0 & \\
0 & 0 & \overline{M}^1 & & \cdots & 0 & \\
\vdots & & & & \ddots & \vdots & \\
0 & 0 & 0 & & \cdots & \overline{M}^p &
\end{pmatrix} \tag{17}$$

Here, the $n \times n$ sub-matrix M_0 has the form

$$M_0 = D^2 f - \sum_{j=1}^{p} \overline{\mu}_j D_x^2 \mathcal{L}^{\overline{y}^j} + \sum_{j=1}^{p} \overline{\mu}_j \big(- D_{yx}\mathcal{L}^{\overline{y}^j} Dy^j + D_x^T v^j D\gamma^j \big)\ . \tag{18}$$

In view of (11) it follows that

$$-D_{yx}\mathcal{L}^{\overline{y}^j} Dy^j = D^T y^j D_y^2 \mathcal{L}^{\overline{y}^j} Dy^j - D^T \gamma^j D_y v^j Dy^j = D^T y^j D_y^2 \mathcal{L}^{\overline{y}^j} Dy^j + D^T \gamma^j D_x v^j\ .$$

By substituting this relation into (18) we find that M_0 equals the matrix $\overline{M}_0$ in (14) with $\overline{\mu}_0 = 1$. In view of our assumptions (7) and (13) the matrix $\left(\begin{smallmatrix} M_0 & -B^T \\ B & 0 \end{smallmatrix}\right)$ is regular. Hence, by using (12), the matrix (17) and therefore also the matrix (16) is regular. $\square$

In practice, to obtain a 'globally convergent' Newton-type method, one has to apply a (globally convergent) method for finite problems to the locally reduced problems $\mathrm{GSIP}_{\mathrm{loc}}(x)$. For SIP such an algorithm is described in [4, Algorithm 7.4]. With the modifications indicated in Section 2 this algorithm can directly be applied to GSIP. Another possibility is to calculate an approximate solution of GSIP by a discretization method as given in the next section and to use this approximation as a starting value for the solution of the system (15) by Newton's method.

4 Discretization methods for GSIP

Another way for solving SIP are discretization methods (see e.g. [1], [4] for a survey). In this section we will generalize these methods from the SIP-case to GSIP. Due to the dependence of the sets Y on x this generalization is not immediate. The difficulties in comparison with the situation for SIP are discussed.

For given compact sets $Y^1, Y^0 \subset I\!\!R^r$ we define the distances

$$d(Y^1, Y^0) = \max_{y^0 \in Y^0} \min_{y^1 \in Y^1} ||y^1 - y^0|| \quad , \quad d_{\mathrm{H}}(Y^1, Y^0) = \max\{d(Y^1, Y^0) \ , \ d(Y^0, Y^1)\} \ .$$

Let us introduce some assumptions.

A1. Given the compact set M^0 in $I\!\!R^n$, the set valued mapping $Y : I\!\!R^n \to 2^{I\!\!R^r}$ satisfies condition (1) and Y is continuous on M^0, i.e. for any $\overline{x} \in M^0$, $\lim_{x \to \overline{x}} d_{\mathrm{H}}(Y(x), Y(\overline{x})) = 0$. ($M^0$ will be fixed later on.)

Remark 1 Condition (1) implies that Y is upper semi-continuous (closed) and that for any $x \in I\!\!R^n$ the set $Y(x)$ is compact such that if $Y(x) \neq \emptyset$, a solution of the lower level problem $Q(x)$ exists. The continuity of Y implies the continuity of the value function $v(x)$ of $Q(x)$. We give a standard result in parametric optimization: Let the following assumption $\mathrm{A}_{\mathrm{MFCQ}}$ be satisfied.

$\mathrm{A}_{\mathrm{MFCQ}}$: Let for all $x \in M^0$ the MFCQ hold for $Q(x)$, i.e. for any $y \in Y(x)$ we have (4).

Then, $Y(x)$ is (Lipschitz-) continuous on M^0 (M^0 compact) in the following sense. There exist $c > 0$ such that

$$d_{\mathrm{H}}(Y(x_1), Y(x_2)) \leq c \, ||x_2 - x_1|| \quad \text{for all} \quad x_1, x_2 \in M^0 \ .$$

For SIP, the assumption A1 simply means that the (fixed) set Y is compact. The following assumption is also standard in SIP.

A2. The feasible set M of GSIP is compact.

This condition implies that a (global) solution of GSIP exists. Let v_{GSIP} denote the minimal value of GSIP, $v_{\mathrm{GSIP}} = \min_{x \in M} f(x)$.

Remark 2 Since the continuity assumption on Y implies that M is closed (cf. [6]), condition A2 can also be replaced by the assumption that M is bounded. This condition can always be imposed by adding constraints $|x_i| \leq \kappa,\ i = 1, \dots, n$ for some large $\kappa > 0$. Note, that for non-continuous mappings Y the set M need not be closed in general (cf. Example 1 below).

A discretization method is based on discretizations of the sets $Y(x)$. In any step of such a method we have to choose discretizations $Y^*(x)$ of $Y(x)$ such that for any x, the set $Y^*(x)$ is a finite set satisfying $Y^*(x) \subset Y(x)$. Then, we solve the problem

$$\text{GSIP}(Y^*): \quad \min\ f(x) \quad \text{subject to } x \in M^* = \{x \in \mathbb{R}^n \mid g(x, y) \geq 0,\ y \in Y^*(x)\} \qquad (19)$$

For SIP, the discretization Y^* is a finite subset of the compact set Y (not depending on x) and thus, GSIP(Y^*) represents a finite optimization problem. However, for GSIP the situation is more complicated. Even under the assumption A1 and A2 the discretization $Y^*(x)$ need not be continuous in x and the feasible set M^* need not be closed (i.e. a solution of GSIP(Y^*) may not exist). We give an illustrative example.

Example 1 Consider the GSIP

$$\max\ x \quad \text{s.t. } x \in M = \{x \in [-1, 1] \mid x - 2y \geq 0,\ y \in Y(x)\},$$

with $Y(x) = \{y \mid -1 \leq y \leq x\}$. Then, $M = \{x \in [-1, 1] \mid x - 2x \geq 0\} = [-1, 0]$. Choosing the discretization $Y^*(x) = Y(x) \cap \mathbb{Z}$ it follows, $Y^*(x) = \{-1\}$ for $x \in [-1, 0)$, $Y^*(x) = \{-1, 0\}$ for $x \in [0, 1)$, $Y^*(x) = \{-1, 0, 1\}$ for $x = 1$. We find $M^* = [-1, 1)$, which is not closed, and a solution of GSIP(Y^*) does not exist.

To avoid such a bad behavior we have to assume that the discretizations $Y^*(x)$ are also continuous.

A3. Let be given the compact set M^0 in $\mathbb{R}^n$. The discretization $Y^*(x) \subset Y(x)$ is defined by continuous functions $y_i^* : M^0 \to \mathbb{R}^r,\ i = 1, \dots, i_*$,

$$Y^*(x) = \{y_i^*(x),\ i = 1, \dots, i_*\}, \quad x \in M^0 .$$

Now, we are going to generalize the discretization method to GSIP.

Algorithm 1 (*Conceptual discretization method*)

Step k: Given a discretization $Y^k(x) \subset Y(x)$

i. Select a (finer) discretization $Y^{k+1}(x)$, $Y^{k+1}(x) \subset Y(x)$ and compute a solution x^{k+1} of GSIP(Y^{k+1}).

ii. Stop, if x^{k+1} is feasible within a fixed accuracy $\alpha > 0$, i.e. $g(x^{k+1}, y) \geq -\alpha,\ y \in Y(x^{k+1})$.
Otherwise, step $k + 1$.

Theorem 2 *Suppose that the assumptions A1, A2 are satisfied. Let the discretizations* $Y^k(x)$ *of* $Y(x)$ *be choosen such that A3 holds for* $Y^k(x)$ *as well as* $Y^0(x) \subset Y^k(x),\ k \in I\!N$. *Let the feasible set* M^0 *of GSIP(*Y^0*) be compact. Suppose,*

$$d(Y^k(x), Y(x)) \to 0 \quad \textit{for} \quad k \to \infty, \quad \textit{uniformly on the (compact) set} \quad M^0. \tag{20}$$

Then, the sequence $\{x^k\}$ *of solutions* x^k *of GSIP(*Y^k*) has an accumulation point* $\overline{x}$ *and each such point is a solution of GSIP.*

Proof. By assumptions A1, A2, A3 and using $Y^0(x) \subset Y^k(x) \subset Y(x)$ the feasible sets M, M^k respectively, of GSIP, GSIP(Y^k) respectively, are compact (cf. Remark 2) and satisfy

$$M \subset M^k \subset M^0, \quad k \in I\!N .$$

Consequently, a solution $x^k \in M^k$ of GSIP(Y^k) exist. Since $x^k \in M^0$, the sequence $\{x^k\}$ has an accumulation point $\overline{x} \in M^0$. Without restriction we can assume $x^k \to \overline{x},\ k \to \infty$. In view of $M \subset M^k$, the values $f(x^k)$ and v_{GSIP} fulfill $f(x^k) \le v_{\text{GSIP}}$ and thus by continuity of f we find

$$f(\overline{x}) \le v_{\text{GSIP}} .$$

It suffice to show that $\overline{x} \in M$. Let $\overline{y} \in Y(\overline{x})$ be given arbitrarily. Since $d(Y(x^k), Y(\overline{x})) \to 0$ for $k \to \infty$ (by continuity of Y) and using (20), we can choose $\hat{y}^k \in Y(x^k)$, $y^k \in Y^k(x^k)$ such that

$$\lim_{k\to\infty} \hat{y}^k = \overline{y} , \quad \lim_{k\to\infty} |y^k - \hat{y}^k| = 0 .$$

In view of $g(x^k, y^k) \ge 0$, by taking the limit $k \to \infty$, it follows $g(\overline{x}, \overline{y}) \ge 0$, i.e. $\overline{x} \in M$.

□

We also generalize the so-called exchange method from SIP to GSIP. This method can be more efficient than a pure discretization method as given in Algorithm 1.

Algorithm 2 (*Conceptual exchange method*)

Step k: Given a discretization $Y^k(x) \subset Y(x)$ and a fixed, small value $\alpha > 0$.

i. Compute a solution x^k of GSIP(Y^k).

ii. Calculate local solutions $y_i^k,\ i = 1, \dots, i_k$ ($i_k \ge 1$) of $Q(x^k)$ (cf. (2)) such that one of them, say y_1^k, is a global solution, i.e. $g(x^k, y_1^k) = \min_{y \in Y(x^k)} g(x^k, y)$

iii. Stop, if $g(x^k, y_1^k) \ge -\alpha$ (solution $\overline{x} \approx x^k$).
Otherwise, construct functions $y_i^k(x)$ continuous on $I\!R^n$ such that $y_i^k(x^k) = y_i^k$, $y_i^k(x) \in Y(x),\ i = 1, \dots, i_k$ and put

$$Y^{k+1}(x) = Y^k(x) \cup \{y_i^k(x),\ i = 1, \dots, i_k\} . \tag{21}$$

To ensure the convergence of this algorithm we have to make a further assumption.

A4. Given the compact set M^0, the functions $y_1^k(x),\ k \in I\!N$, are equicontinuous on M^0, i.e. for any $\varepsilon > 0$ there exists $\delta > 0$ such that $x_1, x_2 \in M^0, |x_1 - x_2| < \delta$ implies $|y_1^k(x_1) - y_1^k(x_2)| < \varepsilon$ for all $k \in I\!N$.

Theorem 3 *Suppose that the assumptions A1, A2 are satisfied. Let $Y^0(x)$ be chosen such that A3 holds and that the feasible set M^0 of GSIP(Y^0) is compact. Let A4 be satisfied. Then, the exchange method in Algorithm 2 with $\alpha = 0$ either stops with a solution $\overline{x} = x^{k_0}$ of GSIP or the sequence $\{x^k\}$ of solutions of GSIP(Y^k) has an accumulation point $\overline{x}$ and each such point is a solution of GSIP.*

Proof. We consider the case that the algorithm does not stop with a solution. As in the proof of Theorem 2, by our assumptions, a solution x^k of GSIP(Y^k) exist and $x^k \in M^0$. Thus, the sequence $\{x^k\}$ has an accumulation point $\overline{x} \in M^0$ and again we can assume $x^k \to \overline{x},\ k \to \infty$. As before we find

$$f(\overline{x}) \leq v_{\text{GSIP}}$$

and we only have to show that $\overline{x} \in M$, i.e. $v(\overline{x}) \geq 0$ for the value function $v(x)$ of $Q(x)$. In view of $v(x^k) = g(x^k, y_1^k)$ (see Algorithm 2ii) we can write

$$v(\overline{x}) = v(x^k) + v(\overline{x}) - v(x^k) = g(x^k, y_1^k) + v(\overline{x}) - v(x^k) \ .$$

Since $y_1^k(x) \in Y^{k+1}(x)$ we have $g(x^{k+1}, y_1^k(x^{k+1})) \geq 0$ and in view of A4 it follows using $y_1^k = y_1^k(x^k)$ that $|y_1^k(x^{k+1}) - y_1^k| \to 0$ for $k \to \infty$. Consequently, by continuity of g and v (cf. Remark 1) we find

$$v(\overline{x}) \geq \big(g(x^k, y_1^k) - g(x^{k+1}, y_1^k(x^{k+1}))\big) + \big(v(\overline{x}) - v(x^k)\big) \to 0 \quad \text{for} \quad k \to \infty \ . \qquad \square$$

After deriving the convergence results we have to discuss how strong the assumptions A1-A4 are. We furthermore indicate how the assumption A3 can be fulfilled in practice.

From the theoretical point of view, the only severe assumption is the condition in A1 that the set-valued mapping Y is continuous. This condition is not fulfilled in general (in the generic case) since in particular it excludes that by changing x a (connected) component of $Y(x)$ may disappear (or a new component may appear). Recall that a sufficient condition for the continuity of Y is the condition A_{MFCQ}. We give an example.

Example 2 Consider the problem

$$\text{P:} \quad \min\ x^2 \quad \text{s.t.} \quad x \in M = \{x \in [-2, 2] \mid y - x - 1 \geq 0,\ y \in Y(x)\},$$

with $Y(x) = \{y \in I\!R \mid 0 \leq y,\ y \leq -x\}$. We find $Y(x) = \left\{ \begin{smallmatrix} [x,0] \ , & x \leq 0 \\ \emptyset \ , & x > 0 \end{smallmatrix} \right.$ and $M = [-2, -1] \cup (0, 2]$. At $\overline{x} = 0$ the condition MFCQ is not fulfilled for $Y(\overline{x}) = \{0\}$. Obviously, the mapping Y is not continuous at $\overline{x} = 0$, M is not closed and a solution of P does not exist.

We now outline a possible way to construct a continuous discretization $Y^*(x)$ of $Y(x)$ as given in A3. In practice, this has only to be done locally near a given point $\overline{x}$ (where the actual computation takes place). Under assumption A1 or the stronger condition A_{MFCQ} such a construction is always possible. Note, that A_{MFCQ} implies that for x near $\overline{x}$, $x, \overline{x} \in M^0$, the sets $Y(x)$ and $Y(\overline{x})$ are (Lipschitz-) homeomorphic (cf. [2, Theorem B]).

We give the construction for the case that $Y(x)$ is a set in $I\!R^2$. Assume that $Y(x) \subset C_0,\ x \in M^0$ (cf. (1)). Let be given $\overline{x}$ of M^0 and an appropriate, small neighborhood $U(\overline{x})$ of $\overline{x}$.

Construction of $Y^*(x)$ in $U(\overline{x})$: Choose a mesh size h and define the grid points $p_{i,j} = h(i,j),\ i,j \in \mathbb{Z}$. Choose $N \in I\!N$ such that $C_0 \subset \{(y_1, y_2) \mid -hN \leq y_i \leq hN,\ i = 1, 2\}$. Initialize index sets, $I_1 = I_2 = \emptyset$, and proceed as follows:

For $i, j = -N$ to N do:

1. If $p_{i,j} \notin Y(\overline{x})$, goto 4, else goto 2.
2. If $p_{i+\rho,j+\tau} \in Y(\overline{x})$ for all $\rho, \tau = -1, 0, 1$ (neighbors of $p_{i,j}$) then put $I_1 = I_1 \cup \{(i,j)\}$, $y^{i,j}(x) = p_{i,j},\ x \in U(\overline{x})$ and goto 4
   else put $I_2 = I_2 \cup \{(i,j)\}$ and goto 3.
3. For $\rho, \tau = -1, 0, 1$ (or some other ordening) do :

   if $p_{i+\rho,j+\tau} \notin Y(\overline{x})$, then put $\rho_i = \rho, \tau_j = \tau$ and define $y^{i,j}(x)$ to be the intersection point of the line $l(t) = p_{i,j} + t(p_{i+\rho_i,j+\tau_j} - p_{i,j})$, $|t|$ minimal, with the boundary $\partial Y(x)$. Goto 4.

   Next ρ, τ.
4. Next i, j.

Then, the desired discretization is given by $Y^*(x) = \{y^{i,j}(x)\ ,\ \ (i,j) \in I_1 \cup I_2\}$.

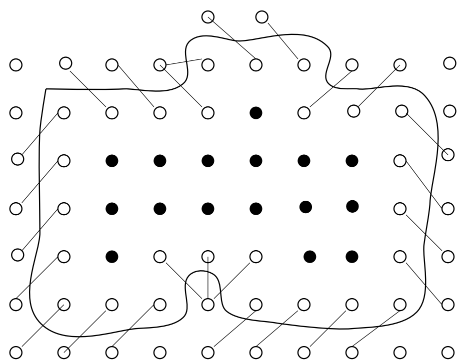

Figure 1 Illustration of the construction of the discretization $Y^*(x)$.

Remark 3 Clearly, the 'size' of the neighborhood $U(\overline{x})$ where the discretization $Y^*(x)$ constructed above can be used, is strongly related to the mesh size h chosen in the construction. The neighborhood $U(\overline{x})$ should necessarily satisfy the condition

$$p_{i,j} \in Y(x), \quad (i,j) \in I_1, \quad \text{for all } x \in U(\overline{x}) \ .$$

In a forthcoming paper [13] we will investigate numerically whether this construction can be implemented in practice such that the convergence of the Algorithms 1 and 2 are not affected.

For SIP, the set Y and the discretization Y^* do not depend on x such that the assumptions A3 and A4 are not relevant. So, one could also try to avoid the construction in A3 by transforming GSIP into a common SIP. In [12] it has been shown that under $\mathrm{A}_{\mathrm{MFCQ}}$ the problem GSIP can be transformed to an equivalent SIP (with functions $\tilde{g}(x,y)$ which need only to be Lipschitz-continuous). However, in the general case this transformation is constructed by locally defined functions which are 'glued together' in an abstract way. Hence, this transformation is only useful if the set valued function Y satisfies certain convexity conditions. See [12, Lemma 1] for such a construction. In [13] numerical experiments will be done.

5 Convex GSIP

In this section we answer the question under which conditions a GSIP is a convex problem, i.e. under which conditions the feasible set of GSIP is convex and the first order condition is sufficient for optimality.

Similar to the situation in finite optimization, the following is true for SIP.

Theorem 4 *Let be given a problem SIP. Suppose, f is convex and for any (fixed) y the function $-g(x,y)$ is convex in x (on $\mathbb{R}^n$). Then we have:*

(a) *The feasible set M of SIP is convex.*

(b) *Suppose, for $\overline{x} \in M$ the Kuhn-Tucker condition is satisfied, i.e. with $\overline{\mu}_0 = 1$, $\overline{\mu}_1, \dots, \overline{\mu}_p \geq 0$ the equation (8) holds. Then, $\overline{x}$ is a (global) minimizer of SIP.*

For GSIP the situation is more complicated due to the dependence of Y on x. This is illustrated by the problem in Example 2 where the feasible set $M = [-2,-1] \cup (0,2]$ is not convex although all problem functions are linear.

We firstly give a sufficient condition for M to be a convex set.

Lemma 2 *Suppose that the function $-g(x,y)$ is convex in (x,y) (on $\mathbb{R}^{n+r}$) and assume that the following set-valued inclusion holds: For any $x_1, x_2 \in \mathbb{R}^n$ and α, $0 < \alpha < 1$ we have,*

$$Y(\alpha x_1 + (1-\alpha)x_2) \subset \alpha Y(x_1) + (1-\alpha)Y(x_2) \,. \tag{22}$$

Then, the feasible set M of GSIP is convex.

Proof. Let be given $x_1, x_2 \in \mathbb{R}^n$, $0 < \alpha < 1$. Put $x_\alpha = \alpha x_1 + (1-\alpha)x_2$. Choose $y \in Y(x_\alpha)$ arbitrarily. In view of (22), there exist $y_1 \in Y(x_1)$, $y_2 \in Y(x_2)$ such that $y = \alpha y_1 + (1-\alpha)y_2$. By convexity of $-g$ we find

$$g(x_\alpha, y) = g(\alpha(x_1,y_1) + (1-\alpha)(x_2,y_2)) \geq \alpha g(x_1,y_1) + (1-\alpha)g(x_2,y_2) \geq 0 \,. \qquad \square$$

To illustrate condition (22) we have depicted in Figure 2 two possible situations for the case that $x, y \in I\!R$.

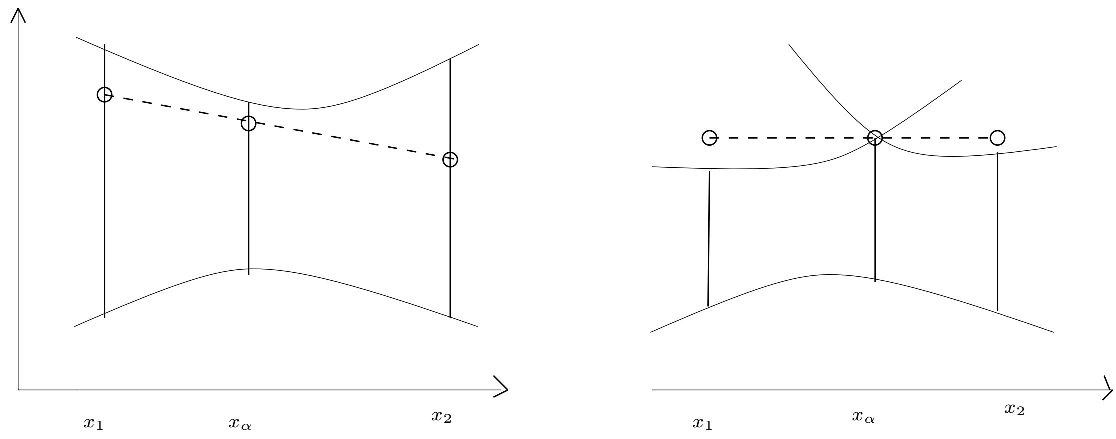

Figure 2 a) Condition (22) is satisfied b) Condition (22) is not satisfied

From Figure 2b it is clear that for $Y(x) = \{y \in I\!R \mid v_l(x,y) \geq 0,\ l \in L\},\ x \in I\!R$ the condition (22) cannot be satisfied if there exist points $\overline{x}, \overline{y} \in Y(\overline{x})$ such that $v_1(\overline{x},\overline{y}) = v_1(\overline{x},\overline{y}) = 0$ and the gradients $Dv_1(\overline{x},\overline{y}), Dv_2(\overline{x},\overline{y})$ are linearly independent. So, roughly speaking, the boundary of the set $\{(x,y) \mid y \in Y(x)\}$ may not have 'corners' as in Figure 2b.

Before proving our main result, we need a lemma.

Lemma 3 *Let be given $\overline{x} \in I\!R^n$ and a point $\overline{y}^1 \in Y(\overline{x})$ such that at $\overline{y}^1$ the condition LICQ holds for $Q(\overline{x})$. Then, there exist a neighborhood $U(\overline{x})$ of $\overline{x}$ and a C^1-function $y^1 : U(\overline{x}) \to I\!R^r$, such that $y^1(\overline{x}) = \overline{y}^1$, $y^1(x) \in Y(x)$ and $v_l(x, y^1(x)) = 0$ for all $l \in L_0(\overline{x}, \overline{y}^1),\ x \in U(\overline{x})$.*

Proof. The result follows by applying the implicit function theorem to the equations $v_l(x,y) = 0,\ l \in L_0(\overline{x},\overline{y}^1)$. □

Theorem 5 *Suppose, the Kuhn-Tucker condition for GSIP is satisfied at $\overline{x} \in M$, i.e. (6) holds with $\overline{\mu}_0 = 1$ and points $\overline{y}^1, \ldots, \overline{y}^p \in Y_0(\overline{x})$. Suppose, the assumptions of Lemma 2 hold and LICQ is satisfied for $Q(\overline{x})$ at all active points $\overline{y}^1, \ldots, \overline{y}^p$. Let furthermore f be convex (in x) and $v_l(x,y),\ l \in L$, be convex in $((x,y))$. Then, $\overline{x}$ is a global minimizer of GSIP.*

Proof. By convexity of $-g$, for any $x \in M, y^j \in Y(x)$ we obtain,

$$0 \leq g(x, y^j) - g(\overline{x}, \overline{y}^j) \leq D_x g(\overline{x}, \overline{y}^j)(x - \overline{x}) + D_y g(\overline{x}, \overline{y}^j)(y^j - \overline{y}^j)\ . \tag{23}$$

Choose now the neighborhood $U(\overline{x})$ and the functions $y^j = y^j(x)$ according to Lemma 3 corresponding to the points $\overline{y}^1, \ldots, \overline{y}^p \in Y_0(\overline{x})$. Using convexity of v_l it follows

$$0 = v_l(x, y^j) - v_l(\overline{x}, \overline{y}^j) \geq D_x v_l(\overline{x}, \overline{y}^j)(x - \overline{x}) + D_y v_l(\overline{x}, \overline{y}^j)(y^j - \overline{y}^j)\ . \tag{24}$$

Thus for any $x \in U(\overline{x}) \cap M$ we find by using the convexity of f, the Kuhn-Tucker condition (6) as well as (10), (23), (24) that

$$
\begin{aligned}
f(x) - f(\overline{x}) &\geq Df(\overline{x})(x - \overline{x}) \\
&= \sum_{j=1}^{p} \overline{\mu}_j \Big(D_x g(\overline{x}, \overline{y}^j)(x - \overline{x}) - \sum_{l \in L_0(\overline{x}, \overline{y}^j)} \overline{\gamma}_l^j D_x v_l(\overline{x}, \overline{y}^j)(x - \overline{x}) \Big) \\
&\geq -\sum_{j=1}^{p} \overline{\mu}_j \Big(D_y g(\overline{x}, \overline{y}^j)(y^j - \overline{y}^j) + \sum_{l \in L_0(\overline{x}, \overline{y}^j)} \overline{\gamma}_l^j D_x v_l(\overline{x}, \overline{y}^j)(x - \overline{x}) \Big) \\
&\geq -\sum_{j=1}^{p} \overline{\mu}_j \sum_{l \in L_0(\overline{x}, \overline{y}^j)} \overline{\gamma}_l^j \Big(D_y v_l(\overline{x}, \overline{y}^j)(y^j - \overline{y}^j) + D_x v_l(\overline{x}, \overline{y}^j)(x - \overline{x}) \Big) \geq 0 \, .
\end{aligned}
$$

Hence, $\overline{x}$ is a local minimizer on $U(\overline{x}) \cap M$. Since f is convex (on the convex set M), $\overline{x}$ is also a global minimizer. □

References

[1] Görner S., Reemtsen R., *Numerical methods for semi-infinite programming: A survey*, in ‘Semi-infinite Programming’, Reemtsen/Rückmann (eds.), Kluwer, Boston (1998).

[2] Guddat J., Jongen H.Th. and Rückmann J., *On stability and stationary points in nonlinear optimization*, J. Australian Math. Soc., Series B 28, 36-56, (1986).

[3] R. Hettich, G. Still, *Semi-infinite programming models in Robotics*, in ’Parametric Optimization and Related Topics II’, Guddat et al. (eds.), Akademie Verlag, Berlin (1991).

[4] Hettich R., Kortanek K., Semi-infinite programming: Theory, methods and applications, SIAM Review, vol 35, No.3, 380-429, (1993).

[5] R. Hettich, G. Still, *Second order optimality conditions for generalized semi-infinite programming problems*, Optimization Vol. 34, pp. 195-211, (1995).

[6] H. Th Jongen, J.-J. Rückmann, O. Stein, *Generalized semi-infinite optimization: A first order optimality condition and examples*, to appear in Mathematical Programming.

[7] H. Th Jongen, J.-J. Rückmann, O. Stein, *Disjunctive optimization: critical point theory*, J. Optim. Theory Appl. 93, no. 2, 321-336 (1997).

[8] Kaplan A. and Tichatschke R., *On the numerical treatment of a class of terminal problems*, Optimization Vol. 41, 1-36, 1997.

[9] Levitin E. and Tichatschke R., *A branch and bound approach for solving a class of generalized semi-infinite programming problems*, J. of Global Optimization 13, 299-315, (1998).

[10] Rückmann J.-J., Shapiro A., On first order optimality conditions in generalized semi-infinite programming, to appear.

[11] Stein O.,*On the level sets of marginal functions*, University of Aachen, Preprint No. 80, (1998).

[12] Still G., *Generalized semi-infinite programming: Theory and methods*, to appear in European Journal of Operational Research.

[13] Still G., *Generalized semi-infinite programming: Computational experiments with numerical methods*, in preparation.